Training algorithms for artificial neural network in predicting of the content of chemical elements in the upper soil layer

Cite as: AIP Conference Proceedings **2048**, 060004 (2018); https://doi.org/10.1063/1.5082119
Published Online: 11 December 2018

Andrey Shichkin, Alexander Buevich, Alexander Sergeev, Elena Baglaeva, Irina Subbotina, Julian Vasilev, and Maria Kehayova-Stoycheva

Training Algorithms for Artificial Neural Network in Predicting of the Content of Chemical Elements in the Upper Soil Layer

Andrey Shichkin[1,2, a)], Alexander Buevich[1,2, b)], Alexander Sergeev[1,2, c)], Elena Baglaeva[1,2, d)], Irina Subbotina[2, e)], Julian Vasilev[3, f)], Maria Kehayova-Stoycheva[3, g)]

[1]*Ural Federal University, Mira str., 19, Ekaterinburg, RUSSIA 620002*
[2]*Institute of Industrial Ecology UB RAS, S.Kovalevskoy str., 20, Ekaterinburg, RUSSIA 620990.*
[3]*University of Economics, 77 Knyaz Boris I Blvd., Varna, Bulgaria 9002*

[a)]and@ecko.uran.ru
[b)]bagalex3@gmail.com
[c)]Corresponding author: aleksandrpsergeev@gmail.com
[d)]elenbaglaeva@gmail.com
[e)]iesub@mail.ru
[f)]vasilev@ue-varna.bg
[g)]kehajova_m@ue-varna.bg

Abstract. Models based on Artificial Neural Networks (ANN) in recent years are increasingly being used in environmental studies. Among the many types of ANN, the network type Multilayer Perceptron (MLP) has become most widespread. Such networks are universal, simple, and suitable for most tasks. The main problem when modelling using MLP is the choice of the learning algorithm. In this paper, we compared several learning algorithms: Levenberg-Marquart (LM), LM with Bayes regularization (BR), gradient descent (GD), and GD with the speed parameter setting (GDA). The data for modelling were taken from the results of the soil screening of an urbanized area. The spatial distribution of the chemical element Chromium (Cr) in the surface layer of the soil was simulated. The structure of the MLP network was chosen using computer simulations based on minimization of the root mean squared error (RMSE). The model using the LM training algorithm showed the best accuracy.

INTRODUCTION

The ANN model might be applied to the raw data obtained in the environmental monitoring, and then can be used to forecast the chemical elements content at the unmonitored locations [1, 6,11].The behavior of chemical elements in soils is complex and their content is depends on many factors [10, 13]: the properties of elements, chemical and physical properties of soils, as well as environmental factors and their effect on soil-geochemical processes.

The most common used ANN in studying of the chemical elements distribution in soil is multilayer perceptrons(MLP) [7, 9,12].In a conventional MLP model, the input neurons represent predicting variables.They are connected to a single or multiple layer(s) of hidden neurons, which are then linked to the output neurons representing the target variable.The construction of the MLP model is the selection of parameters: the number of hidden layers and the number of neurons within each hidden layer. During the ANN training process, the connections between the neurons are established by assigning weights based on an intrinsic learning process where the weights are iteratively adjusted to match the outputs of the training dataset [2, 7]. MLP model uses the

Proceedings of the 44th International Conference on Applications of Mathematics in Engineering and Economics
AIP Conf. Proc. 2048, 060004-1–060004-5; https://doi.org/10.1063/1.5082119
Published by AIP Publishing. 978-0-7354-1774-8/$30.00

iterativeprocedure of learning choosesa learning algorithm.There are many different training algorithms, with different characteristics and performance.

Any learning algorithmfor neural networks has both advantages and shortcomings [2, 12]: 1) learning algorithmapproaches are oftenheuristic; 2) the problems of the training set preparation are related to the difficulty of finding a sufficient number of learning examples; 3)no universal learning algorithm.

Inthe present studyMLP model with different learning algorithms for the spatial forecasting of element (Chromium) content is usedto predict the element concentration in topsoil. Raw data of the element content was obtained in topsoil at a particular location of Tarko Salecity, Russia.

MATERIALS AND METHODS

Raw data for the study were obtained from the results of the soil survey in *Tarko Sale, south part of the residual area*, *Yamalo-Nenets Autonomous Okrug*, Russia (N64.9°, E77.8°). This region is a subarctic climatic area (Köppen climate classification Dfc). Tarko Sale soils were formed in Pyaku-Pur river drainage areas. The local soils are the podzols with a little content of organic matter [3]. The composition of the soil is 100% sand (size fraction less than 1 mm).In total, 101 topsoil samples at a depth of 0.05 m were taken from inhabited zones of the city insites of undamaged natural soil. The detailed spatial location of sampling points is shown in Fig. 1.Chromium contents in soil samplesfor MLP modelwere obtained by a chemical analysis.Preparation of soil specimens and chemical analysis were conducted in compliance with actual standard requirementsof the Russian federal certification system. The chemical laboratory meets the general requirements for the competence of testing and calibration laboratories ISO/IEC 17025:2005.

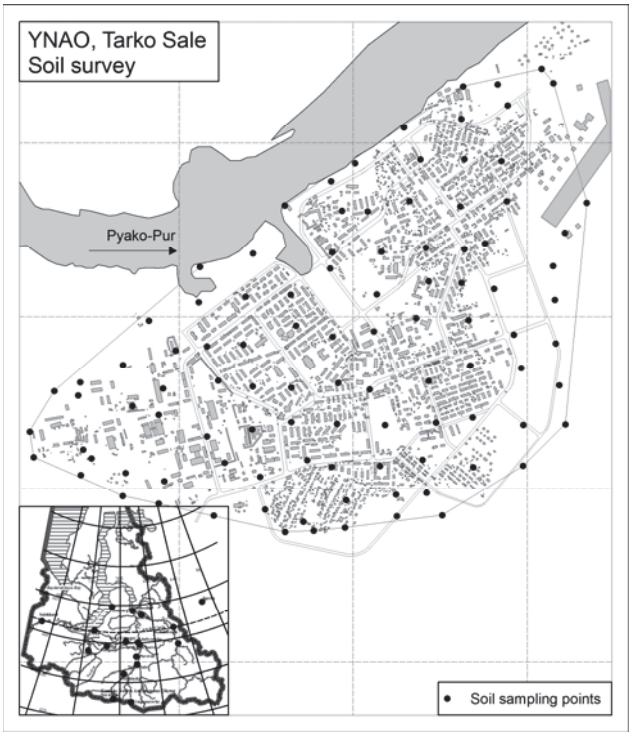

FIGURE1. The sampling area*Tarko Salecity*, *Yamalo-Nenets Autonomous Okrug*, *Russia*

The environmental raw data wereCr content u in i-location with coordinate $\mathbf{s_i}=\{x_i, y_i\}$. These measured contents are realizations of the spatial field $c(\mathbf{s})$. The descriptive statistics were calculated by STATISTICA.

The structure of ANN is feed-forward MLP with different training algorithms [2]: Levenberg-Marquart (LM), LM with Bayes regularization (BR), gradient descent (GD), and GD with the speed parameter setting (GDA). All the sampling points were randomly split into independent training and test data sets. The input layer of MLP was coordinates of sampling points of the training set (70% sampling points); the hidden layer consisted of several neurons, and the output layer represented the element content. The selection of the number of neurons in the hidden

layer was carried out by the lower total RMSE of prediction of the element content. The number of neurons was varied from 1 to 20. Each network was trained 500 repetitions and the best of them was selected. Network learning quality was checked by RMSE between the results of the network predictions and the test data set. The test data set (30% sampling points) was used for testing the models only.

The learning problem for MLP is formulated as a procedure for modifying the weight and biases of a network. In order to move the outputs $u(\boldsymbol{s})$ closer to the targets $\mathbf{s_i}$.The objective is to reduce the mean square error, meaning of which is the differencebetween the neuron response $u_{ANN}(\boldsymbol{s})$ and the target vector $u(\boldsymbol{s})$ at which the loss function (mean square error) takes a minimum value.

GD is the simplest training algorithm with fixed value of the training rate. It trains any network as long as its weight, net input, and transferfunctions have derivative functions.Backpropagation is used to calculate derivatives of performance withrespect to the weight and bias variables.Training stops when the maximum number of repetitions is reached.

GDA is a network training algorithm that updates weight and bias values according to gradient descent with adaptive training rate. Unlike the constant training rate in GD, here the training rate is corrected in each repetition by some adaptive coefficient, which dependent on performance.

LM algorithm updates weight and bias values accordingto least-squares method.The loss functions is a sum of squared errorsof the model in the training sample is used as an optimization criterion. The algorithm consists in successive approximation of the given initial values of the parameters to the desired local optimum.Training set is divided on validation and test vectors. Validation vectors are used to stop training early if the network performance on the validation vectors fails to improve or remains the same for maximum number of repetitions. Test vectors are used as a further check that the network is generalizing well, but do not have any effect on training. Training stops when any of these conditions occurs: the maximum number of repetitions is reached, the performance gradient falls below some fix value, validation performance has increased more than a predetermined number of times since thelast time it decreased (when using validation).

BR algorithm is similar to the optimization of Levenberg-Marquardt. It adds minimization of weights to a minimum of error squares. This process is called Bayes regularization. The Bayesian regularization minimizes the linear combination of error squares and weights, so that at the end of the learning the resulting network has good generalizing qualities. Training stops like in LM.

The model rapidity is estimated on a real time spent on computer training.

The predictive accuracy of MLP model with different learning algorithm was verified by the Spearman's rank correlation coefficient and RMSE (1) between the prediction and raw data from the test data set.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(u_{ANN}(\mathbf{s}_i)-u(\boldsymbol{s_i}))^2}{n}}, \quad (1)$$

where $u_{ANN}(\mathbf{s}_i)$is a predicted content (ANNs, kriging), $u(\mathbf{s_i})$ is a measured content, n is a number of points.

RESULTS AND DISCUSSION

The descriptive statistics of *Cr*contentare shown in the Table 1. The element contents are erratic and positively skewed in nature. Chromium demonstrates an extremely high maximum content of 1424 mg/kg, which reflected at two spots, while median value is only 86.9 mg/kg.Comparing to both background contents in the Ural Region (Ural Clarke) and in the world soils (World Clarke), the total Cr content at urban background does not exceed the reference values, while the total Cr at anomaly sites was a few times higher than Ural Clarke [8, 14]. Total Cr contents in podzols are known to fall into the range from 2.6 to 34 mg/kg in Canada [4] and from 3 to 200 mg/kg in the USA [10].

TABLE 1. Descriptive statistics of the modeled elements

Number of specimens	Content, mg/kg					CV, %	Skewness	Kurtosis	World Clarke*, mg/kg	Average content (Podzols)**, mg/kg
	Min	Max	Mean	SD	Median					
101	35	1424	259	337	87	130	1.6	1.2	200	2.6-34.0

* *World Clarke* -average content in soils of the world [3].

** Element content is in podzols [14].

K-S test indicates that Cr content distribution is not normal distribution. Cr contents show a bimodal distribution at Tarko Sale (see Table 1). Thus, used for predicting the chromium content data are not simple to build the model of the element distribution in the top soil.

MLP framework selection is based on RMSE minimization: root mean square error (RMSE) of the neural network for testdata depends of different neuron number in the hidden layer for different learning algorithms (see Fig. 2a). The model rapidity for different learning algorithms hangs upon neuron number in the hidden layer (see Fig. 2b).

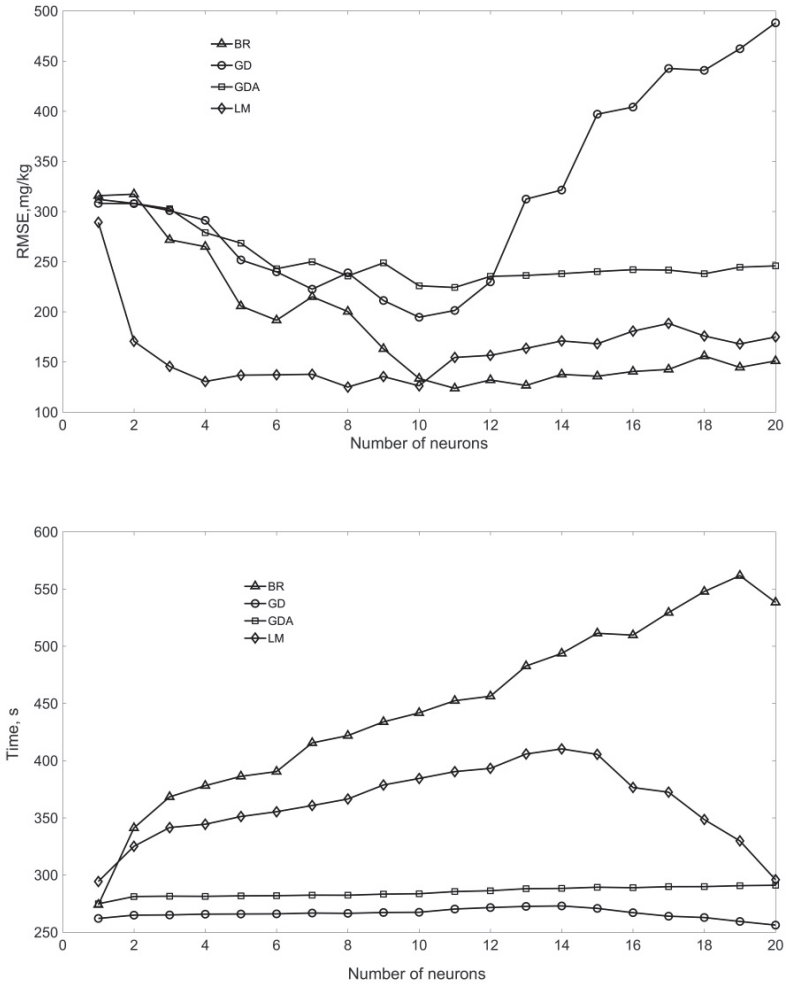

FIGURE2. MLP framework selection based on RMSE minimization (a): root mean square error (RMSE) of the neural network and model rapidity (b)under different neuron number in the hidden layer for different learning algorithms

The final configurations and the accuracy assessment of the MLP model with RMSE for the test data set are presented in Table 2. The hidden layers consisted from 8 to 11 neurons.The shortest training time for GD and GDA algorithms is 267 and 286 seconds at 10 and 11 hidden neurons, respectively (see Table 2). The Spearman's rank correlation coefficient is 0.84 for GD (2-10-1 configuration) and 0.80 for GDA (2-11-1 configuration). In this case the RMSE between the predicted and raw data is 75% and 86% of the mean (259 mg/kg, see Table 1), respectively.

The least RMSE (124 mg/kg) and 11 hidden neurons are demonstrated by the BR algorithm. Despite the RMSE for LM algorithm is 125 mg/kg, the number of neurons in the hidden layer is only 8, training time is 366 s and the Spearman's rank correlation coefficient is highest (0.94, see Table 2).

TABLE 2. Accuracy assessment of MLP model indices of the *Cr* concentration

Learning algorithm	Number of the test data set	MLP configuration	Spearman's rank correlation coefficient	RMSE, mg/kg	Time, s
GD	31	2-10-1	0.84	195	267
GDA	31	2-11-1	0.80	224	286
LM	**31**	**2-8-1**	**0.94**	**125**	**366**
BR	31	2-11-1	0.93	124	452

Table 2shows the parameters used to compare the performance of the different methods.The best results demonstrated by MLP with LM algorithm(are in **bold**).

The MLP model using the LM training algorithm showed the best accuracy to the forecasting of the chemical elements distribution in the top soil with comparison of different training algorithms for MLP model.

CONCLUSION

Thus, comparison of different learning algorithms for MLP model to the prediction of the chemical elements distribution in the surface layer of soil was carried. The best learning algorithm was MLP model using the LM training algorithm.Comparative analysis was performed on data of chromiumsoil distribution in Tarko Sale, YNAO, Russia.The computer training timeand the root-mean-square error (RMSE)were applied as the efficiency indicators of the models for the validation (test) data set.

REFERENCES

[1] T. Behrens, H. Förster, T. Scholten, U. Steinrücken, E.-D. Spies, M. Goldschmitt, "Digital soilmapping using artificial neural networks", Journal of Plant Nutrition and Soil Science, Vol. 168, 21–33 (2005).

[2] H. Demuth, M. Beale, "Neural Network Toolbox For Use with MATLAB", The Math Works, Inc (2002)

[3] G.V. Dobrovolsky, I.S. Urusevskaya, "Geography of soils", MSU Press, Moscow [in Russian] (2004)

[4] R. Frank, K. Ishida, P. Suda,"Metals in agricultural soils of Ontario", Canadian Journal of Soil Science, Vol. 56, 181-196 (1976).

[5] A. Kabata-Pendias, "Trace elements in soils and plants", Taylor and Francis GroupCRC Press, 201-260 (2011).

[6] M. Kanevski, A. Pozdnoukhov, V. Timonin,"Machine Learning for Spatial Environmental Data.Theory, Applications and Software", EPFL Press (2009)

[7] M. Leuenberger, M. Kanevski,"Extreme Learning Machines for spatial environmental data", Computers & Geosciences, Vol. 85, Part B, 64-73 (2015).

[8] A.P. Sergeev, E.M. Baglaeva, A.V. Shichkin "Case of soil surface chromium anomaly of a northern urban territory – preliminary results", Atmospheric Pollution Research, Vol. 1, 44-49 (2010).

[9] A.P.Sergeev, A.G. Buevich, A.N. Medvedev, I.E. Subbotina, M.V. Sergeeva,"Artificial neural network and kriging interpolation for the chemical elements contents in the surface layer of soil on a background area",*SGEM International Multidisciplinary Scientific GeoConference, Book 3 Vol. 2.*(Bulgaria, 18-24June 2015), pp. 27–34.

[10] H.T. Shacklette, J.G. Boerngen, "Element concentrations in soils and other surficial materials of the conterminous United States", U.S. Geological Survey professional paper, Vol. 1270, 105. (1984).

[11] R.R. Shaker, T.J. Ehlinger,"Exploring non-linear relationships between landscape and aquatic ecological condition in southern Wisconsin: A GWR and ANN approach", International Journal of Applied Geospatial Research, Vol. 5(4), 1-20 (2014).

[12] A.J. Shepherd, "Second-Order Methods for Neural Networks: Fast and Reliable Training Methods for Multi-Layer Perceptrons", Springer-Verlag (1997)

[13] D.A. Tarasov, A.G. Buevich, A.P. Sergeev, A.V. Shichkin,"High Variation Topsoil Pollution Forecasting in the Russian Subarctic: Using Artificial Neural Networks Combined with Residual Kriging", Applied Geochemistry, Vol. 88, Part B, 188-197(2018).

[14] Y.N. Vodyanitskii,"Standards for the contents of heavy metals and metalloids in soils",Eurasian Soil Science, Vol. 45, N. 3, 321–328 (2012).